Building a High Quality and Comprehensive Tandem Mass Spectral Library

Xiaoyu Yang, Pedatsur Neta, Yamil Simón-Manso, Lisa Kilpatrick, Yuxue Liang, Stephen E. Stein

National Institute of Standards and Technology, Gaithersburg, MD

National Institute of Standards and Technology
U.S. Department of Commerce

Abstract

Authentic samples were analyzed on low and high resolution, ion-trap and linear collision cell mass spectrometers. A clustering algorithm used an adjusted dot product as a measure of spectral similarity to create a 'consensus spectrum' from multiple spectra for the same precursor. Each consensus spectrum was examined based on mass accuracy and fragments. The library contains 7,002 compounds, 15,183 precursor ions, 121,591 spectra of positive and negative ions at different collision energies.

Background

- Mass spectral library searching is an effective method for chemical identification because of its reliable searching results and fast searching speed (e.g. metabolite identifications in Standard Reference Materials studies);
- The quality, size, and completeness of a library are the key to successful searching.
- We aim to provide the largest possible collection of high quality reference spectra of biologically and environmentally relevant compounds, many of which are human metabolites.

Procedure of building MS/MS library

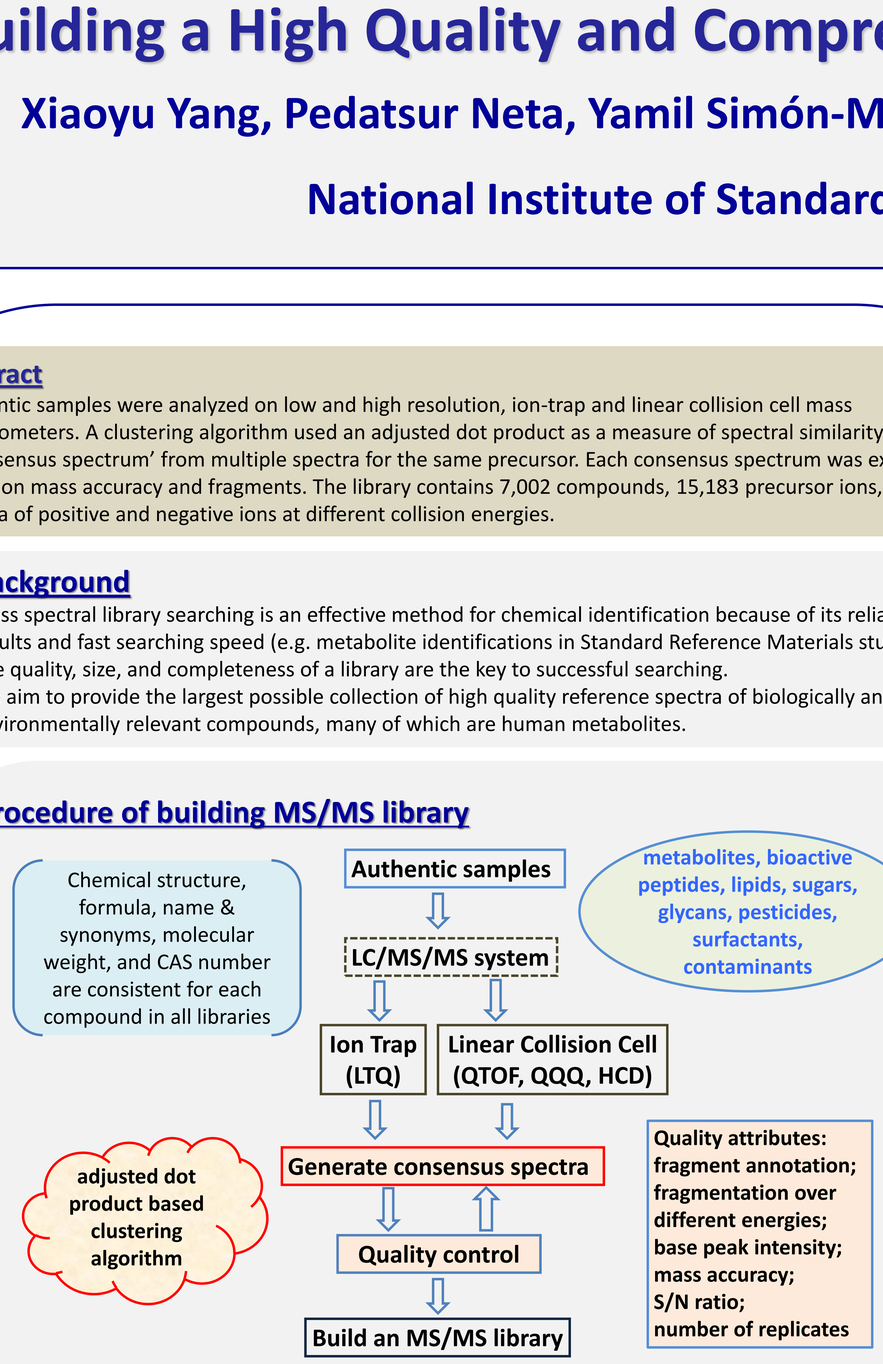


Fig. 1. Flow chart of building MS/MS library

Example 1 - Making and selecting consensus spectra

- An adjusted dot product based clustering algorithm was used to group similar spectra into the same cluster and created one consensus spectrum from each cluster; the best consensus spectrum was picked for the library, whilst other low quality, impurity or contaminant spectra were eliminated.
- Noise peaks were removed using a voting algorithm.
- Example: two clusters were generated for Palatinose [M+NH4]+ (Fig. 2).
  The spectrum in Fig. 2A was kept in the library.

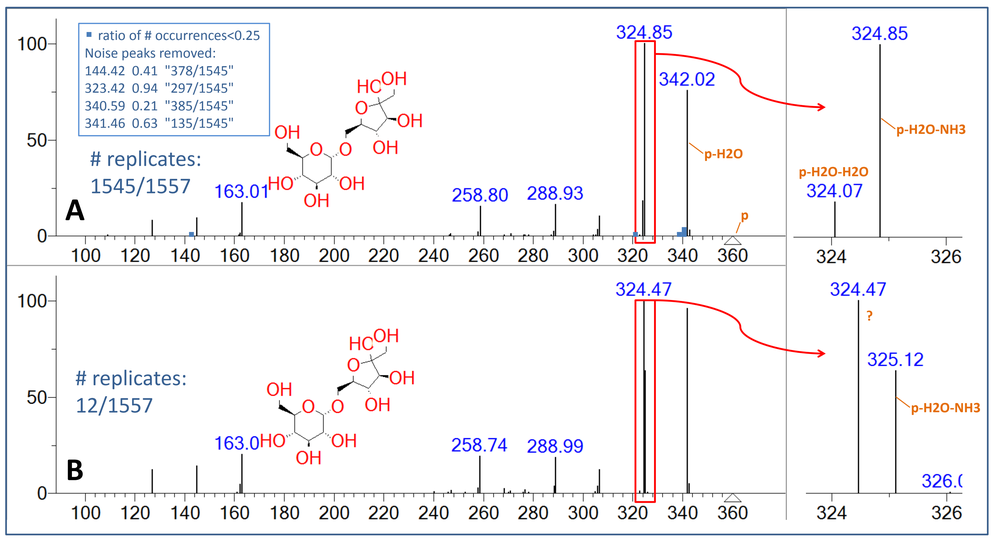


Fig. 2. Generating consensus spectra for Palatinose [M+NH4]+ on LTQ

Example 2 - Peak annotation: Each spectrum was ascertained that all major peaks are assigned to acceptable fragmentation product ions from the known precursor structure (Fig. 3).

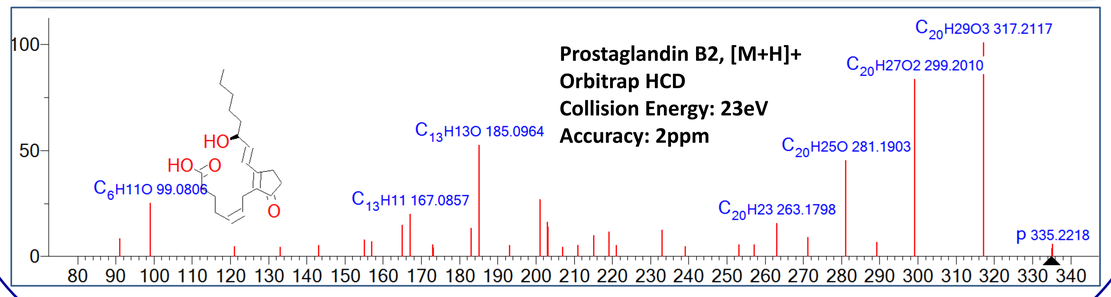


Fig. 3. Annotating spectrum peaks based on precursor structure

Example 3 - Fragmentation at different energies: Each spectrum was ascertained that peak intensities vary with collision energy in a reasonable progression (Fig. 4).

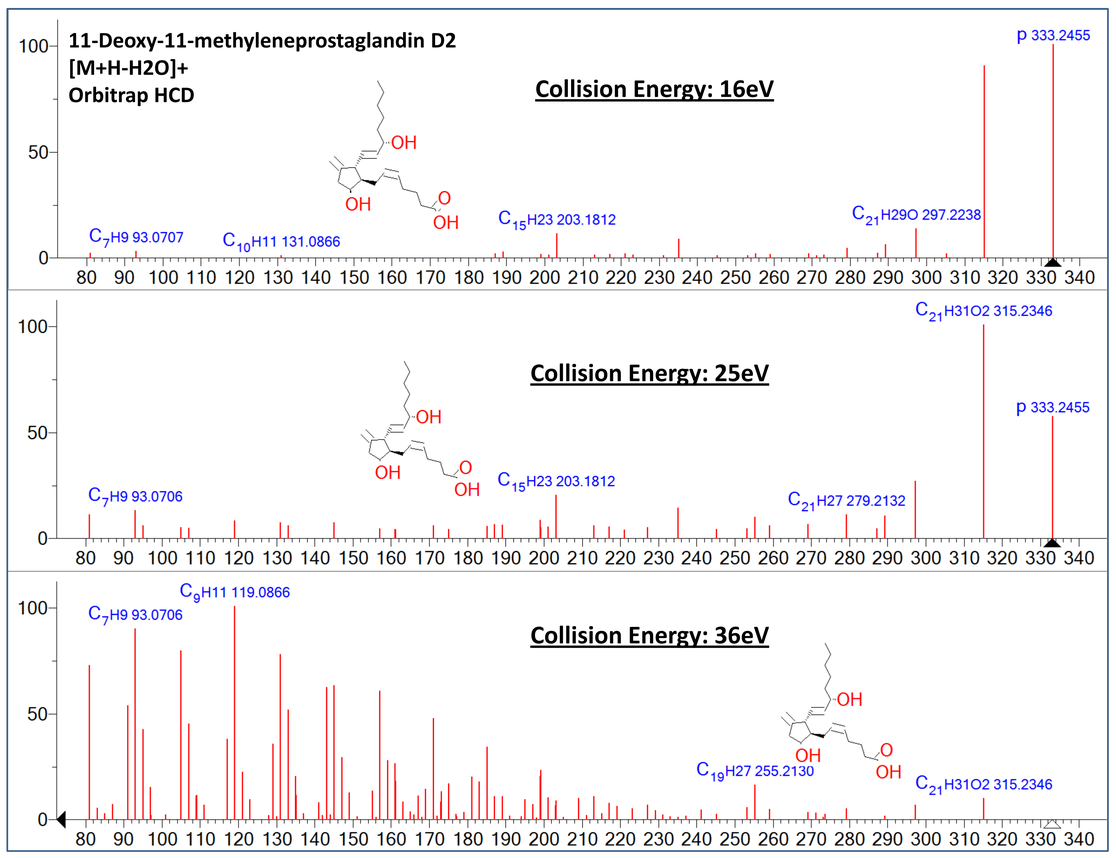


Fig. 4. Dependence of spectrum on collision energy

Example 4 - Different instruments, different precursor types, etc.: The same sample was run on the different instruments (Fig. 5) at positive or negative modes with various precursor types. The fragmentation patterns were compared for confirmation and to give users more accurate searching results.

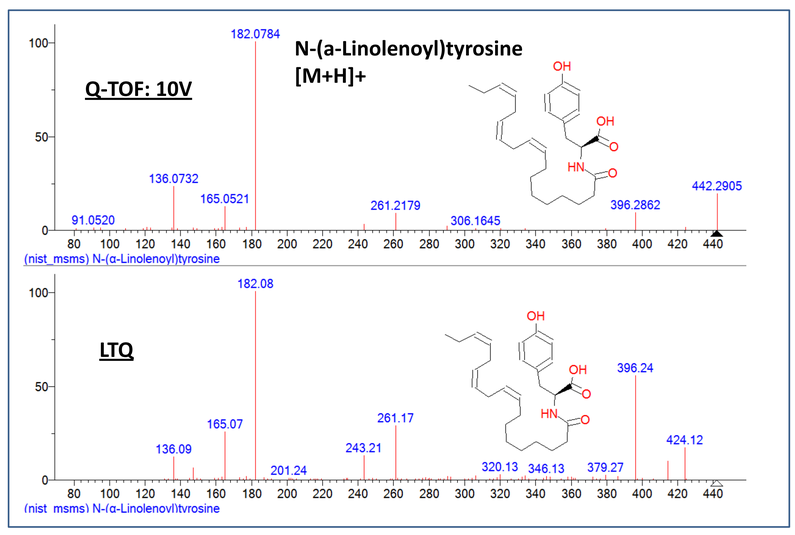
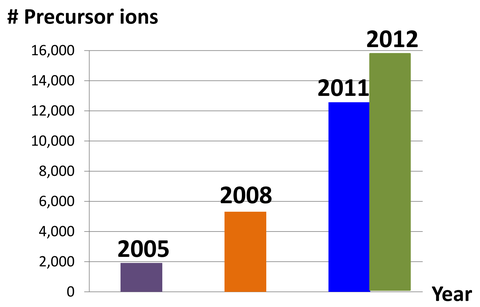


Fig. 5. N-(a-Linolenoyl)tyrosine was analyzed on QTOF and LTQ instruments

Conclusions

- A high quality, comprehensive library is being developed for metabolites, peptides, lipids, sugars, glycans, pesticides, surfactants, and contaminants, etc.
- The MS/MS library will be valuable in metabolomics.



7,002 Compounds
15,183 Precursor Ions
121,591 Spectra
90% Positive Ion Spectra
10% Negative Ion Spectra

Instrument Type	Precursor Ions
Ion Trap	12,049
Collision Cell (QTOF, QQQ, HCD)	9,233

Fig. 6. MS/MS library development

Major precursor types:
$[M+H]^+$, $[M+2H]^{2+}$, [M-H]-, $[M+Na]^+$, $[M+NH4]^+$, $[Cat]^+$, [An]-, [p-H2O], [p-NH3]